Supplemental Figures 1-6

Compares White males with. . .

	. . . *White females*	. . . *Black males*	. . . *Black females*	
Proposal 1 (high quality) is non-White-male	**List 1** *Proposal 1, White female* Proposal 2, White male Proposal 3, White male	**List 2** *Proposal 1, Black male* Proposal 2, White male Proposal 3, White male	**List 3** *Proposal 1, Black female* Proposal 2, White male Proposal 3, White male	Four proposals define a set. Two of the proposals (1-2) are high quality and two of the proposals (3-4) are moderate quality. Each proposal is manipulated into White male, White female, Black male, and Black female versions. The proposal versions are used to create 12 lists, which together test experimental comparisons between White males and each of the three other social categories.
Proposal 2 (high quality) is non-White-male	**List 4** Proposal 1, White male *Proposal 2, White female* Proposal 4, White male	**List 5** Proposal 1, White male *Proposal 2, Black male* Proposal 4, White male	**List 6** Proposal 1, White male *Proposal 2, Black female* Proposal 4, White male	
Proposal 3 (moderate quality) is non-White-male	**List 7** Proposal 1, White male *Proposal 3, White female* Proposal 4, White male	**List 8** Proposal 1, White male *Proposal 3, Black male* Proposal 4, White male	**List 9** Proposal 1, White male *Proposal 3, Black female* Proposal 4, White male	
Proposal 4 (moderate quality) is non-White-male	**List 10** Proposal 2, White male Proposal 3, White male *Proposal 4, White female*	**List 11** Proposal 2, White male Proposal 3, White male *Proposal 4, Black male*	**List 12** Proposal 2, White male Proposal 3, White male *Proposal 4, Black female*	Each list of proposals is sent to 3 reviewers, yielding 36 reviewers per set of proposals. There are 12 sets of proposals, yielding 432 reviewers.

Fig. S1. A set of proposals and proposal versions, which are used to obtain the reviews from a cohort of 36 reviewers. Moderate quality proposals are shown in red, high quality proposals in blue, and, within each list, the proposal that has a non-White-male PI is italicized.

Overall Impact or Criterion Strength	Score	Descriptor
High	1	Exceptional
	2	Outstanding
	3	Excellent
Medium	4	Very Good
	5	Good
	6	Satisfactory
Low	7	Fair
	8	Marginal
	9	Poor

Fig. S2. NIH scoring criteria.

The reviewers were also asked to evaluate additional special considerations, if applicable, including human subjects considerations, protections for vertebrate animals, biohazards, resource sharing plans for multiple PI proposals, and the budget and period of support. Finally, the reviewers were asked to provide an overall verbal evaluation and Overall Impact score. At NIH, this Overall Impact score is typically given the greatest weight during the discussion of reviews and the assignment of a Priority Score (which is used to determine funding lines).
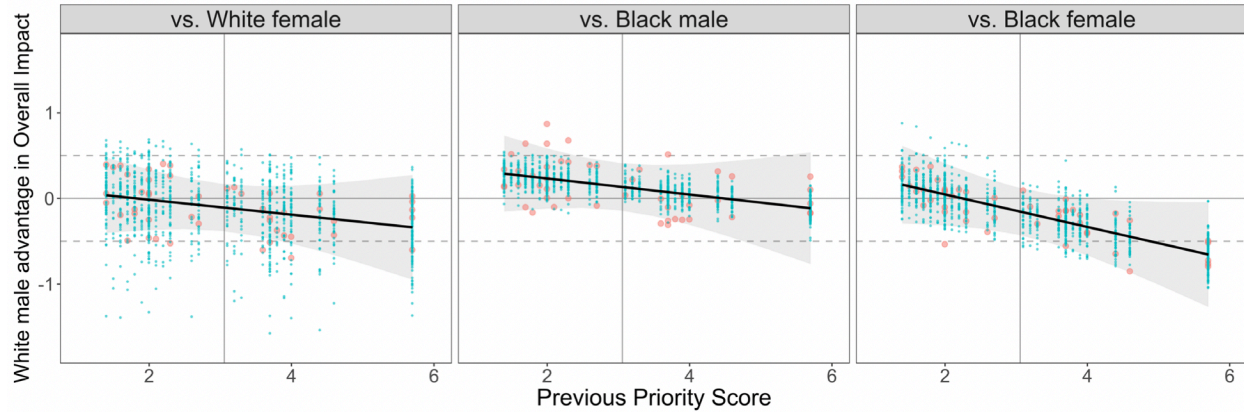
Fig. S3. Relationship between proposal quality and the difference in the Overall Impact scores attained by White male vs non-White-male PIs. The confidence band is a Wald 95% CI; blue dots are reviewer-level random effects, pink dots are grant-level random effects. Quality is operationalized by the Priority Scores given to the original proposals, and the vertical grey line is the mean Priority Score across the 48 proposals. Although descriptively, Black female PIs have an advantage on low quality proposals relative to White male PIs, the overall relationship between Priority Scores and the White male vs. Black female Overall Impact difference is not different from 0.
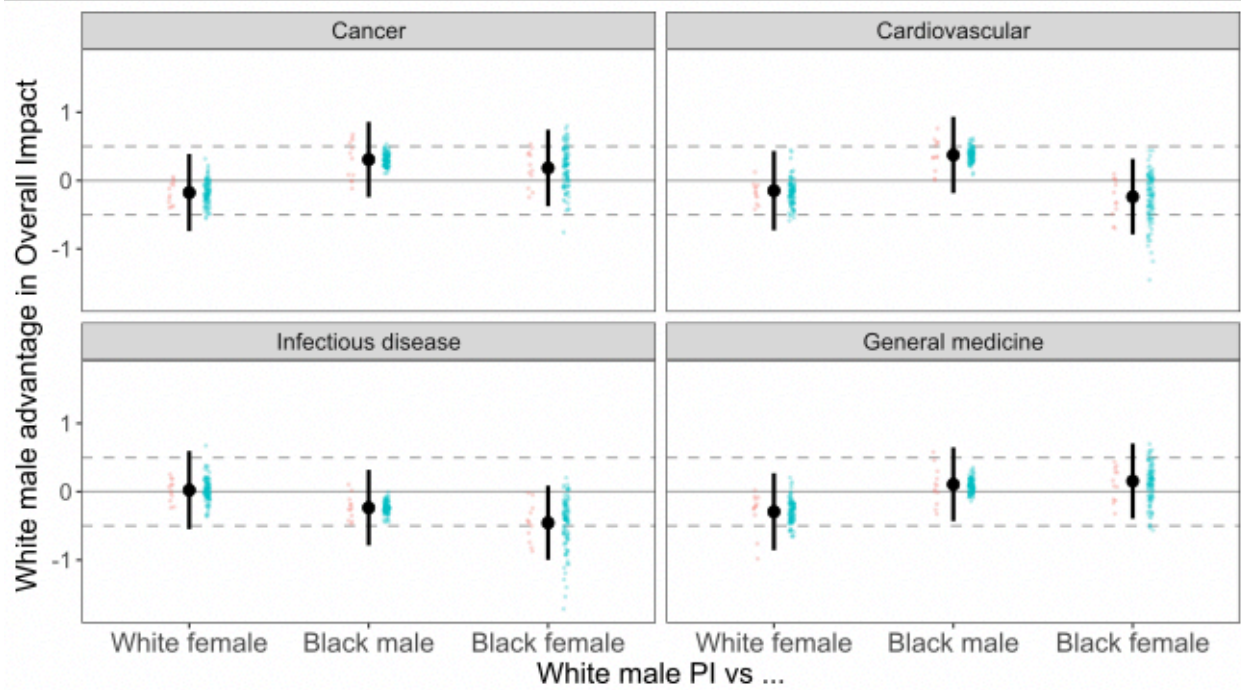
Fig. S4. The difference in the Overall Impact scores attained by White male vs non-White-male PIs in four broad topic areas of science. Topics are defined by the NIH institute that originally funded the proposals assigned to a given reviewer. "Cancer" is the National Cancer Institute; "Cardiovascular" the National Heart, Lungs, and Blood Institute, "General medicine" the National Institute for General Medical Sciences, "Infectious disease" the National Institute for Allergy and Infectious Disease. Dots are the estimated differences from the LMEM, lines are Wald 95% CIs, points to the left and right of each dot are by-proposal and by-reviewer random effects, respectively.
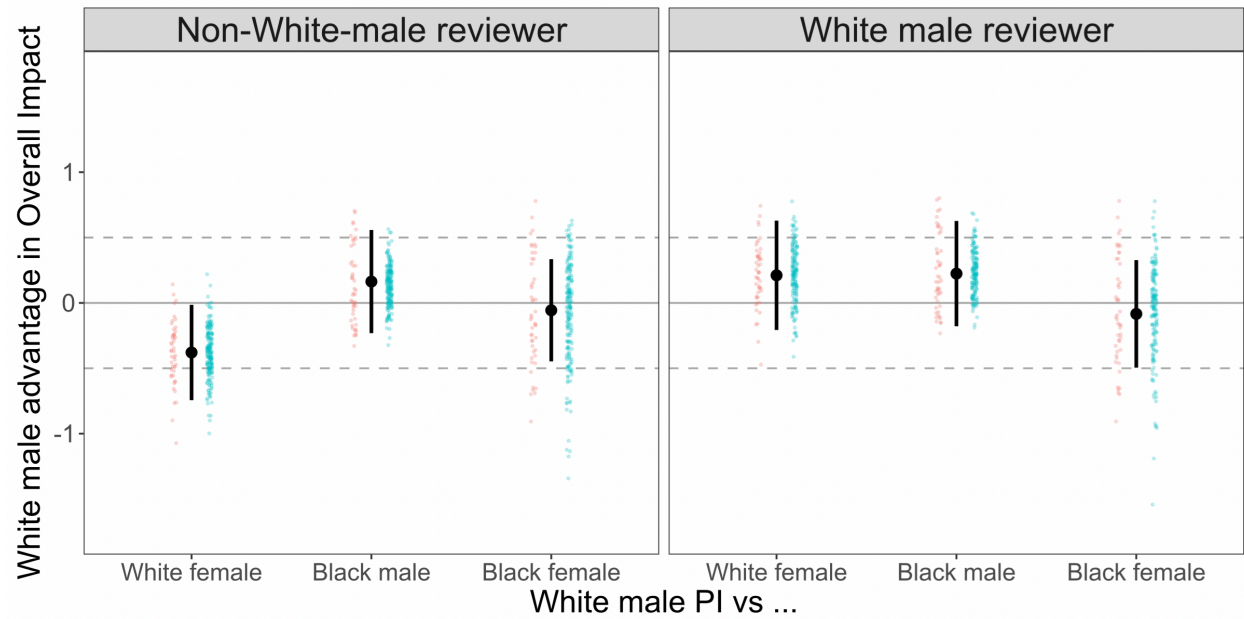
Fig S5. The difference in the Overall Impact scores attained by White male vs non-White-male PIs for non-White- male and White male reviewers. Dots are the estimated differences from the LMEM, lines are Wald 95% CIs, points to the left and right of each dot are by-proposal and by-reviewer random effects, respectively
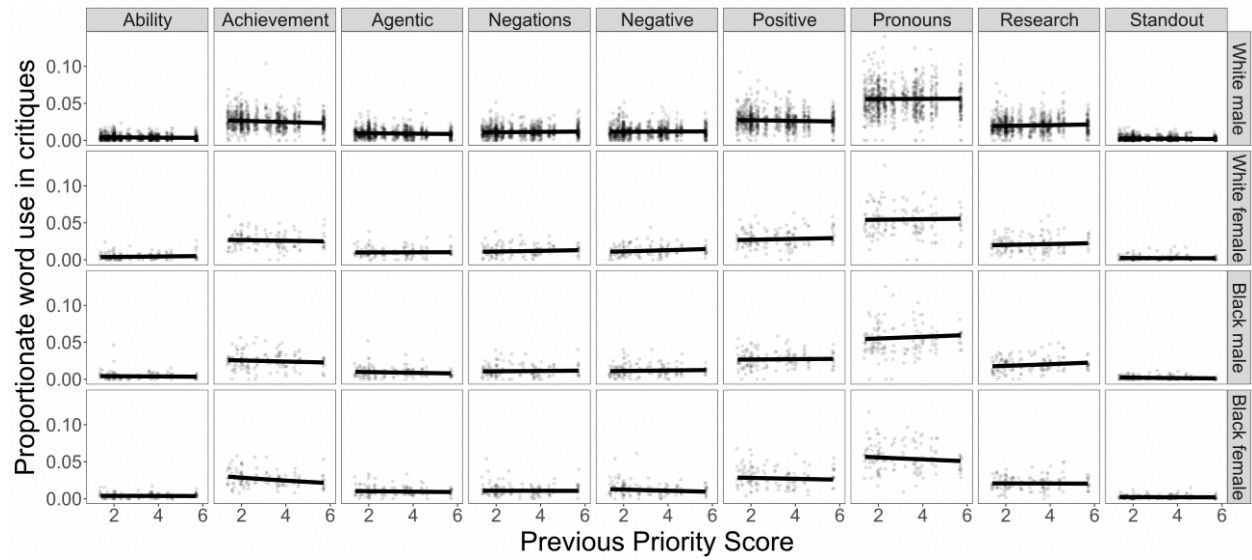
Fig. S6. The relationships between quality, as operationalized by a proposal's previous Priority Score, and the proportionate word use in each of nine categories. Points are jittered to avoid overplotting.